Cloning, Sequencing, and Species Relatedness of the Escherichia coli cca Gene Encoding the Enzyme tRNA Nucleotidyltransferase*

(Received for publication, October 28, 1985)

Henryk Cudny, James R. Lupski$&$, G. Nigel Godson$, and Murray P. Deutscher†

From the Department of Biochemistry, University of Connecticut Health Center, Farmington, Connecticut 06032 and the †Biochemistry Department, New York University Medical Center, New York, New York 10016

The Escherichia coli cca gene which encodes the enzyme tRNA nucleotidyltransferase has been cloned by taking advantage of its proximity to the previously cloned dnaG locus. A series of recombinant bacteriophages, spanning the chromosomal region between the dnaG and cca genes, was subcloned into pBR328, and cells containing the recombinant phage DNA library using recombinant plasmids containing regions between dnaG and cca as probes. Two of the recombinant phage isolates, λc1 and λc4, contained the cca gene. A BamHI fragment from λc1 was subcloned into pBR328, and cells containing this recombinant plasmid, pH9, expressed tRNA nucleotidyltransferase activity at about 10-fold higher level than the wild type control. The cca gene was further localized to a 1.4-kilobase stretch of DNA by Bal31 deletion analysis. The nucleotide sequence of the cca gene was determined by the dideoxy method, and revealed an open reading frame extending for a total of 412 codons from an initiator GTG codon that would encode a protein of about 47,000 daltons. Southern analysis using genomic blots demonstrated that the cca gene is present as a single copy on the E. coli chromosome and that there is no homology on the DNA level between the E. coli cca gene, and the corresponding gene in the Bacillus subtilis, Saccharomyces cerevisiae, Petunia hybrida, or Homo sapiens genomes. Homology was found only with DNA from the closely related species, Salmonella typhimurium. These studies have also allowed exact placement of the cca gene on the E. coli genetic map, and have shown that it is transcribed in a clockwise direction.

The 3' terminal trinucleotide sequence, CCA, is present on all tRNAs and is required for the acceptor and transfer functions of this nucleic acid. Incorporation of AMP and CMP residues into tRNAs that contain incomplete CCA termini is catalyzed by tRNA nucleotidyltransferase (EC 2.7.7.25) (1). The enzyme carries out synthesis of the CCA terminus without the direction of a template using the multiple accepting and donating subsites within its active site (2-4). However, it is still not fully understood how tRNA nucleotidyltransferase recognizes incomplete tRNA molecules, how it synthesizes a defined 3'-end sequence, and which residues on the protein interact with the various substrates.

Genetic studies of tRNA nucleotidyltransferase have only been carried out in Escherichia coli. In this organism tRNA nucleotidyltransferase is encoded by the cca gene which is present at 66 min on the genetic linkage map (5) between the rpsU-dnaG-rpoD macromolecular synthesis operon (6, 7) and the tolC gene (8, 9). Mutants in the cca gene that display altered levels of tRNA nucleotidyltransferase activity have been isolated (10), one of which displays greatly decreased AMP incorporation but normal CMP incorporation (11). Strains carrying a cca mutation accumulate defective tRNA and grow slowly (10). The slow growth phenotype can be overcome by a mutation in relA (12). However, nothing is known about whether the cca gene itself is regulated.

Although tRNA nucleotidyltransferase has been isolated and purified from several species (1), complete analysis of the enzyme has been difficult since the protein is present in cells in relatively low amounts. As a first step for studying the structure and regulation of tRNA nucleotidyltransferase, we have cloned and sequenced the E. coli cca gene. In addition, we have been able to overproduce tRNA nucleotidyltransferase in quantities sufficient for detailed structural and functional characterization (see accompanying paper, Ref. 13).

MATERIALS AND METHODS

RESULTS

CCA Locus of E. coli—The cca gene and surrounding DNA sequences were isolated by directional walking along the E. coli chromosome beginning at the previously cloned dnaG gene (8, 28). The cca gene was first localized to a 4-kb BamHI fragment within this region by assaying for elevated tRNA nucleotidyltransferase activity, and more precisely localized by Bal31 deletion analysis. A restriction cleavage map of this region of the E. coli chromosome is presented in Fig. 1. Based on this restriction map, the distance between the cca and dnaG genes was calculated to be approximately 20 kb, in good agreement with the co-transduction frequency of the genes (35). In addition, the restriction analysis allowed us to

* This work was supported by Grant GM46317 (to M. P. D.) and National Institutes of Allergy and Infectious Disease Grant 7-1142-996 (to G. N. G.). This is paper 46 in the series, "Reactions at the 3' Terminus of tRNA." The previous paper in this series is Ref. 49. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

$ Recipient of a National Institutes of Health Medical Scientist-in-Training Grant.

To whom correspondence should be addressed.

† Portions of this paper (including "Materials and Methods," part of "Results," Figs. 2 and 3, and Tables 1 and 2) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 85M-3595, cite the authors, and include a check or money order for $3.20 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

‡ The abbreviations used are: kb, kilobase(s); bp, base pair(s).
determine the orientation of the 4 kb BamHI fragment relative to dnaG (Figs. 1 and 3), which lead us to conclude that the cca gene is transcribed in a clockwise direction on the E. coli chromosome (Figs. 1 and 3). Nucleotide Sequence Determination—A Bal31 deletion fragment of 2260 nucleotides was sequenced. The restriction map and sequencing strategy are shown in Figs. 2 and 3. Close to 90% of the sequences were determined on both strands, including all of that within the coding region of the cca gene. The nucleotide sequence of the cca gene and surrounding regions is presented in Fig. 4 together with the amino acid sequence predicted for the coding region. Examination of the nucleotide sequence revealed only this single open reading frame beginning at a GTG initiation codon at bp 450, and continuing until a TGA termination codon at bp 1686. This sequence of 1236 bp would encode a 412-amino acid protein of 46,408 daltons. A potential ribosome-binding site and possible −10 and −35 promoter regions were also identified (Fig. 4). The DNA sequenced also includes the 449 nucleotides immediately preceding the cca gene as far as the BamHI site. Within this sequence is part of a potential open reading frame that ends with the terminator codon TAA 63 nucleotides upstream from the cca GTG initiation codon. Examination of the 570-bp nucleotide sequence downstream of the cca gene revealed no sequence that resembled a ρ-independent termination signal (36).

The accuracy of the nucleotide sequence was supported by determination of the amino acid composition of homogeneous tRNA nucleotidyltransferase, isolated as described in the accompanying paper (13). As shown in Table 3, the experimentally determined amino acid composition agrees very well with the predicted composition derived from the DNA sequence.

The cca Gene Is Present as a Single Copy on the E. coli Chromosome and Is Not Homologous to DNA from Other Species—To determine the number of copies of the cca gene on the E. coli chromosome, genomic Southern blots were probed with nick-translated plasmid pRH9 DNA. Total E. coli genomic DNA was individually digested with several restriction enzymes (BamHI, AuaI, KpnI), and as shown in Fig. 5, lanes 2–4, strong hybridization occurs to only one band in each of the digestions. Thus we conclude that the cca gene is present as a single copy on the E. coli chromosome. Similar experiments performed using genomic DNA from Homo sapiens, Bacillus subtilis, and Petunia hybrida revealed no hybridization of these DNAs to the pRH9 probe (Fig. 5, lanes 6–8). Likewise, no hybridization of the pRH9 probe to digested Saccharomyces cerevisiae DNA was observed, even under less stringent conditions of hybridization (6 × SSC, 1 × Denhardt buffer, 42 °C for 12 h) (data not shown). Only the closely related bacterial species Salmonella typhimurium possessed DNA which cross-hybridized to the E. coli cca gene (Fig. 5, lane 5), and in this case the size of the restriction fragment differed from that of E. coli, implying that the sequence around the Salmonella cca gene is not identical to that of E. coli.
which is in good agreement with the genetic mapping data. Analysis of the restriction map obtained from the region cca gene maps between about 15% with experiments have shown that enzyme analysis of cloned into a physical distance of about 11-22 kb site are (35). This places the BamHI fragment. The predicted amino acid sequence of the E. coli chromosomal DNA placed approximately 20 kb from the dnaG gene (5) and surrounding tRNA nucleotidytransferase is also shown. Possible -10 and -35 sequences of a potential promoter region and a possible ribosome-binding site are underlined.

**DISCUSSION**

The cca gene and surrounding E. coli chromosomal DNA sequences have been cloned by directional walking using the adjacent previously cloned dnaG gene as a probe (28). The cca gene maps between 66 and 67 min on the E. coli chromosomal genetic linkage map between the tolC gene and the rpsU-dnaG-rpoD macromolecular synthesis operon (8). Previous experiments have shown that dnaG and rpoD co-transduce about 15% with tolC (37), and cca has been shown to co-transduce with tolC about 50% (5) and with dnaG about 70% (35). This places the cca gene approximately one-quarter to one-half of a minute from the dnaG locus which translates into a physical distance of about 11-22 kb (8). Restriction enzyme analysis of cloned E. coli chromosomal DNA placed cca approximately 20 kb from the dnaG gene (see Fig. 1), which is in good agreement with the genetic mapping data. Analysis of the restriction map obtained from the region around tolC (9) reveals no correlation with the map surrounding cca. In addition, a nick-translated tolC clone does not hybridize to any of the recombinant phages containing E. coli chromosomal DNA sequences between cca and dnaG. Therefore, the tolC gene is not contained within this cloned region of the E. coli chromosome, which is in accord with the published genetic data. With the newly described A. Churon 28 E. coli chromosomal DNA recombinants reported in this paper, and those from a previous paper (28), approximately 30 kb of contiguous E. coli chromosomal sequences have now been physically characterized, including all the genetic loci known to map there: cca, rpsU, dnaG, and rpoD (Fig. 1).

The size of the cca gene-coding region is 1236 bp, sufficient to encode a protein of 412 amino acids with a molecular mass of 46,408 daltons. This value is within range of that estimated by other workers from mobility in sodium dodecyl sulfate-polyacrylamide gel electrophoresis, which has varied from

Fig. 5. Genomic Southern blots using pRH9 as a probe. Total chromosomal DNA isolated from E. coli, S. Typhimurium, B. subtilis, H. sapiens, and P. hybida were digested with various restriction enzymes, and run overnight at 5 V/cm on a 0.8% agarose gel. DNA was transferred to nitrocellulose paper by the method of Southern (23), and nick-translated plasmid pRH9 was then used to probe the genomic blots using the conditions of Ozaki et al. (24). Lane 1, pRH9 DNA, BamHI; lane 2, E. coli DNA, BamHI; lane 3, E. coli DNA, AvoI; lane 4, E. coli DNA, KpnI; lane 5, S. typhimurium DNA, AvoI; lane 6, H. sapiens DNA, HindIII; lane 7, B. subtilis DNA, HindIII; lane 8, P. hybida DNA, HindIII.

Table 3

<table>
<thead>
<tr>
<th>Amino acid residue</th>
<th>Predicted from nucleotide sequence</th>
<th>Determined from protein hydrolysate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>39</td>
<td>42</td>
</tr>
<tr>
<td>Arginine</td>
<td>36</td>
<td>37</td>
</tr>
<tr>
<td>Aspartic acid + asparagine</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>Cysteine</td>
<td>5</td>
<td>5*</td>
</tr>
<tr>
<td>Glutamic acid + glutamine</td>
<td>47</td>
<td>48</td>
</tr>
<tr>
<td>Glycine</td>
<td>28</td>
<td>30</td>
</tr>
<tr>
<td>Histidine</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Leucine</td>
<td>46</td>
<td>46</td>
</tr>
<tr>
<td>Lysine</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Methionine</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Proline</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>Serine</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Throneine</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Valine</td>
<td>32</td>
<td>31</td>
</tr>
</tbody>
</table>

*a* Determined by performic acid hydrolysis.

*b* Determined spectroscopically.

A homology score of 53.8, within the range of known E. coli promoters. Further examination of the possible promoter function of this region is given in the accompanying paper (13). We have also noted the sequence Gly-X-Gly-X-Gly at two positions in the amino acid sequence (beginning at Gly 68 and Gly 272). This sequence has been implicated in nucleotide-binding sites of a number of nucleotide-binding proteins (43).

Codon usage for the cca gene is given in Table 4. Rare codons, as defined by Konigsberg and Godson (44), were found to occur as 5.6% of the total codons in the reading frame. Likewise, optimal codon usage, calculated as described by McLachlan et al. (45), gave a value of 64.6%. Both of these measurements are indicative of a relatively poorly expressed gene, and are consistent with the low level of tRNA nucleotidyltransferase found in E. coli (~200 molecules/cell).

Acknowledgment—We thank Dr. Zhu Liuquin for assistance with the DNA sequence analysis.

REFERENCES

Cloning and Sequencing of E. coli cca Gene


Supplementary Material

Cloning, Sequencing and Species Relationships of the Coding Region of the E. coli cca Gene Encoding the E. coli DNA HelicodiodeTransfers.

Hemphyl Cowley, James E. Lupski, G. Nippos Hoc, and Murray P. Deutscher

METHODS AND MATERIALS

Media and reagents. LB broth and media were as described by Miller (14). LB broth media consists of 1% tryptone, 0.5% NaCl, 0.2% maltose, 0.005% thymine, 0.05% glucose, 0.005% adipic acid, 0.01% yeast extract, and 0.01% vitamin solution (11). Antibiotics obtained from Sigma, are as described in Rigby & Howe, 1984. Antibiotics were obtained from Sigma, and Lasaline, Inc.

Bacterial strains, plasmids, and phages. See Table 1. The E. coli strain 71.25 was constructed by transformation of strain 71.18 (15) with plasmid pBR322. The plasmid DNA was isolated from bacteria grown under shortwave ultraviolet light using the method described by Cohen et al. (27) or that of Lupski et al. (28).

Restriction enzyme analysis. All restriction enzyme analyses were performed on New England Biolabs restriction endonucleases. Electrophoresis of restriction fragments was performed on horizontal agarose slab gels in buffer containing 45 mM Tris-borate, 0.5 M NaCl, 0.5% agarose. Gels were stained with ethidium bromide. Cells were grown in 0.5% agarose plates (Mulligan et al., 1983). Restriction enzyme analysis was performed with New England Biolabs restriction endonucleases. Electrophoresis of restriction fragments was performed on horizontal agarose slab gels in buffer containing 45 mM Tris-borate, 0.5 M NaCl, 0.5% agarose. Gels were stained with ethidium bromide. Cells were grown under shortwave ultraviolet light using a red filter and Polyform Ammonium 05 or 0.5% TMM. Worms were as described by Cohen et al. (27).

Limitations. 44 DNA fragments were collected by centrifugation of 4 liter of medium containing 0.5 M NaCl, 1.2% agarose. 44 DNA bands were collected by centrifugation of 4 liter of medium containing 0.5 M NaCl, 1.2% agarose. DNA fragments were as described by Cohen et al. (27). The concentration of DNA in the ligase mixture was adjusted to favor homologous or intermolecular and products as described (28). Transformation was carried out using the procedure described by Cohen et al. (27) or that of Lupski et al. (28).

Table 1: Bacterial strains, plasmids, and phages.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Species</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>E. coli</td>
<td>Reference 1</td>
</tr>
<tr>
<td>JM109</td>
<td>E. coli</td>
<td>Reference 2</td>
</tr>
<tr>
<td>DH5a</td>
<td>E. coli</td>
<td>Reference 3</td>
</tr>
<tr>
<td>pUC119</td>
<td>Plasmid</td>
<td>Reference 4</td>
</tr>
<tr>
<td>pUC19</td>
<td>Plasmid</td>
<td>Reference 5</td>
</tr>
</tbody>
</table>

Assays of DNA HelicodiodeTransfers. Cells were grown in minimal media to an A 0.5 for 4 hr and 1.0 hr of growth. 4 ml of culture were collected by centrifugation of 4 liter of medium containing 0.5 M NaCl, 1.2% agarose. DNA fragments were collected by centrifugation of 4 liter of medium containing 0.5 M NaCl, 1.2% agarose. DNA bands were collected by centrifugation of 4 liter of medium containing 0.5 M NaCl, 1.2% agarose. DNA fragments were as described by Cohen et al. (27). The concentration of DNA in the ligase mixture was adjusted to favor homologous or intermolecular and products as described (28). Transformation was carried out using the procedure described by Cohen et al. (27) or that of Lupski et al. (28).
Cloning and Sequencing of E. coli cca Gene

DNA sequence analysis. Sequences were determined by the dideoxy chain termination method (49) using the M13 phage and lambda phage system (33) and a synthetic 15 base universal primer (50). Clones for sequencing were generated by ligation of fragments obtained by restriction enzyme digestion or by exonuclease Bal31 deletions. Templates for sequencing were prepared as described in Sanger et al. (52).

DNA and amino acid analysis. Protein for amino acid analysis was dialyzed extensively against 20 mM tris(hydroxy)aminomethane (Tris) buffer, pH 7.5, and hydrolyzed with 6 N HCl for 24 or 72 h at 107°C in evacuated tubes. The equivalent of 1 mg of protein was analyzed on an updated Beckman 121 amino acid analyzer.

RESULTS

Cloning of the cca gene and surrounding E. coli chromosomal DNA sequences.

A lambda phase 26 plaque library containing partial Sau3AI EcoRI fragments of E. coli DNA was used as the source of overlapping, lambda fragments. The library was first probed with labeled PstI EN fragments which contain DNA plus 5 kb of upstream DNA, i.e., on the 5' side of the cca gene (Fig. 1). Nine Sau3AI EcoRI fragments which hybridized to the probe were isolated from over 10,000 plaques that were screened. These nine plaque recombinants were then probed with a second plasmid, pB344, containing only the cca gene (20), in order to remove plaques containing overlapping fragments from outside the cca gene. Three plaque recombinants were identified that hybridized strongly to PstI EN DNA weakly or not at all to pB344. They were designated pB36, pB71, and pB29, and presumably contained chromosomal DNA in the direction of the cca gene. To test whether the cca gene was present in these recombinants, E. coli cells infected with each of the clones were assayed for tRNA nucleotidytransferase activity. Cells containing pB36 and pB29 expressed E. coli nucleotidytransferase activity above basal levels. Cells containing pB71 were a 20-fold higher level of E. coli nucleotidytransferase activity than a plasmid-free E. coli strain (Table 2). The pB36 region of the cca gene was present on these recombinant plagues.

In order to construct a plasmid carrying the cca gene, DNA from the recombinant phage, PstI EN, was digested with the restriction endonuclease BamHI and the digest ligated into BamHI-cleaved pBR322 (46). The recombinant plasmids were used to transform competent E. coli HB101 cells following the procedure of Cohen et al. (22). Ampicillin-resistant, tRNA nucleotidytransferase-sensitive colonies were isolated and assayed for E. coli nucleotidytransferase activity. These cells with elevated activity were found to harbor a recombinant pBR322 plasmid containing a 4.0-kb EcoRI fragment from pB36. This recombinant pBR322 plasmid was designated pRH9. Cells harboring pRH9 show a 20-fold higher level of E. coli nucleotidytransferase activity than a plasmid-free E. coli strain (Table 2).

Table 2

<table>
<thead>
<tr>
<th>Phage recombinant activity</th>
<th>Transformation extract</th>
<th>Overproduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRPH9</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>pRB7</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>pRB1</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>pRBM13</td>
<td>0.25</td>
<td></td>
</tr>
</tbody>
</table>

E. coli cells (5 x 10^6) were grown in T1 liquid broth supplemented with 0.2% galactose to an OD600 of 0.9. Cultures were centrifuged at 4000 g for 10 min, and the pellets suspended in 1 ml of 100 mM tris(HCl), pH 7.5. Cells were washed twice and resuspended in 50 mM tris(HCl), pH 7.5, and the suspension adjusted to 10^8 cells/ml. The 10-ml suspension was sonicated for 2 min (600 W, 180 cycles, 2 sec on and 2 sec off). The resulting suspension was heated to 50°C for 10 min and then chilled on ice. The sonicated extract was clarified by centrifugation at 10,000 g for 10 min.

Evaluation of the cca gene by Bal31 deletion analysis.

Sequence analysis of the cca gene was further analyzed to determine the extent of the cca gene within the 4.0-kb BamHI fragment of pRH9. Several fragments with deletions from each end of the BamHI fragment were constructed, and tested for their ability to elevate E. coli nucleotidytransferase activity (Fig. 2). The sites of all deletions were determined by restriction enzyme mapping and by DNA sequencing using single stranded DNA. As shown in Fig. 2, deletion from the artificially designated left end of the BamHI fragment of pRH9 (the end with the PstI site closest to the cca gene in BsmAI) was inserted for only a 3.0-kb distance before cca gene product activity is lost. In contrast, construction F31-2, in which about 400 nucleotides were removed from the left end, results in CCA activity in the pRH9 DNA for only a 3-kb distance before cca gene product activity is lost. This deletion analysis of pRH9 shows that the 3.2-kb region is about 1.2 kb in length which is sufficient to encode a protein with a molecular weight of approximately 40,000.

Figure 7. Bal31 deletion analysis of recombinant plasmid pRH9.

As in the case of the figure, a partial restriction map of pRH9. The bold face indicates the presumed position of the cca structural gene within pRH9. Under pRH9 are several restriction enzyme recognition sites which were determined by subcloning the appropriate fragments into pBR322 (pB36, pB71, pB11-6, Ph11-10). E. coli nucleotidytransferase activity was assayed in transformed HB101 cells as described in METHODS. The base pair location of the right end of the cca gene within pRH9 is approximately 3.7 kb from the left and BamHI site of pRH9. To the right are noted whether or not cells harboring various clones displayed elevated E. coli nucleotidytransferase activity.

Figure 8. Restriction map of cca gene and sequencing strategy.

The cca structural gene begins at position 305 and terminates at position 1689. Each arrow shows the beginning, and end, and direction of the fragment used for sequencing.